

# Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection



# Prediction of Genetic Contributions and Generation Intervals in Populations With Overlapping Generations Under Selection

**Piter Bijma* and John A. Woolliams†**

**Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen Agricultural University, 6700AH Wageningen, The Netherlands and †Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, United Kingdom*

Manuscript received March 16, 1998
Accepted for publication November 17, 1998

ABSTRACT

A method to predict long-term genetic contributions of ancestors to future generations is studied in detail for a population with overlapping generations under mass or sib index selection. An existing method provides insight into the mechanisms determining the flow of genes through selected populations, and takes account of selection by modeling the long-term genetic contribution as a linear regression on breeding value. Total genetic contributions of age classes are modeled using a modified gene flow approach and long-term predictions are obtained assuming equilibrium genetic parameters. Generation interval was defined as the time in which genetic contributions sum to unity, which is equal to the turnover time of genes. Accurate predictions of long-term genetic contributions of individual animals, as well as total contributions of age classes were obtained. Due to selection, offspring of young parents had an above-average breeding value. Long-term genetic contributions of youngest age classes were therefore higher than expected from the age class distribution of parents, and generation interval was shorter than the average age of parents at birth of their offspring. Due to an increased selective advantage of offspring of young parents, generation interval decreased with increasing heritability and selection intensity. The method was compared to conventional gene flow and showed more accurate predictions of long-term genetic contributions.

MOST natural and artificial populations have overlapping generations. When generations overlap, the generation interval differs from the cohort interval. In quantitative genetics, generation intervals are generally defined as the average age of parents at birth of their offspring. In this definition, generation interval is based on the contributions of parental age classes to newborn offspring; *i.e.*, the average age of parents is calculated as the sum of ages at birth of offspring weighted by the contribution of each age class to newborn offspring. This approach is adopted in the well-known gene flow procedure (Hill 1974). However, if selective advantage (*e.g.*, breeding value) is partly inherited, selection in subsequent generations may affect the genetic contribution of parental age classes to future generations. Thus there may be a difference between generation interval based on contributions to newborn offspring, and generation interval based on contributions to future generations. It has been suggested, therefore, to calculate generation intervals on the basis of selected offspring only (Bichard *et al.* 1973). However, contributions of ancestors to future generations may still deviate from contributions to selected offspring.

Recently, Woolliams *et al.* (1999) found significant differences between generation interval calculated as the average age of parents at the time of birth of a cohort of offspring and generation interval based on the concept of long-term genetic contributions. The latter concept was first introduced by James and McBride (1958) and developed further for the prediction of inbreeding by Wray and Thompson (1990) and Woolliams *et al.* (1993). Predictions for more advanced selection systems, however, resulted in complicated expressions (Wray *et al.* 1994) due to the recursive nature of the prediction procedure. Working on the infinitesimal model (Fisher 1918), Woolliams *et al.* (1999) obtained a simple closed-form approximation for the prediction of long-term genetic contributions by considering Bulmer's (1971) equilibrium genetic parameters, which makes a recursive algorithm redundant. The method of Woolliams *et al.* (1999) covers both discrete and overlapping generations and is applicable to mass selection, index selection, and best linear unbiased prediction selection.

The aim of the current article is twofold. First, two methods of Woolliams *et al.* (1999) for the prediction of long-term genetic contributions in populations with overlapping generations are studied in detail. They illustrate mechanisms that determine the development of pedigree, the contribution of different categories to the genetic makeup of the population in the long term, and the turnover time of genes. The dependency of long-term genetic contributions and generation intervals on

*Corresponding author:* P. Bijma, Animal Breeding and Genetics Group, Department of Animal Sciences, Wageningen Agricultural University, P.O. Box 338, Marijkeweg 40, 6700AH Wageningen, The Netherlands. E-mail: piter.bijma@alg.vf.wau.nl

selective advantage is illustrated in populations with overlapping generations under mass or sib index selection, assuming the infinitesimal model (Fisher 1918).

Second, predictions based on the methods of Woolliams *et al.* (1999) are compared to predictions of long-term genetic contributions and generation intervals based on contributions to unselected newborn offspring, as obtained from conventional gene flow (Hill 1974). Both methods are compared to results obtained from simulated data. Accurate predictions of long-term genetic contributions are an important step toward the prediction of rates of inbreeding in selected populations (Woolliams 1998). The current article focuses on the prediction of genetic contributions and generation intervals; the prediction of rates of inbreeding is in a subsequent article. To show the power of theory of Woolliams *et al.* (1999), predictions of genetic gain based on long-term genetic contributions are also presented, but this is not the main item, as accurate predictions of genetic gain are already well established (*e.g.*, Villanueva *et al.* 1993).

## METHODS

Here we first describe the population structure that was used. Subsequently we describe the concept of long-term genetic contributions and the method of Woolliams *et al.* (1999) for the prediction of long-term genetic contributions in populations with overlapping generations, followed by a description of the relationship between generation interval and genetic contributions. Finally, we describe iterative deterministic and stochastic methods to estimate parameters that are needed to predict long-term genetic contributions and related parameters.

**Population model:** This section describes the genetic model, population structure, and selection strategy for which predictions of genetic contributions were made. The trait considered was assumed to be determined by the infinitesimal model (Fisher 1918). Phenotypic values ($P$) were the sum of additive genetic values ($A$, breeding values) and environmental values ($E$), *i.e.*, $P = A + E$. The population consisted of overlapping generations, and selection was based upon a sib index for a single trait. With parents up to a maximum of $c_{max}$ of age there are $2c_{max}$ categories, one for each sex and age of parent. Categories are indexed by $k$ or by $l$, so $k = 1 \ldots c_{max}$ are males, and $k = c_{max} + 1 \ldots 2c_{max}$ are females. Let age($k$) denote the age of category $k$ [so age(1) = 1 = age($c_{max} + 1$)] and let $n_k$ be the number of parents selected from category $k$. The total number of male and female parents equalled $N_m = \sum_{k=1}^{c_{max}} n_k$ and $N_f = \sum_{k=c_{max}+1}^{2c_{max}} n_k$, respectively. Using random mating, each sire was mated to $d$ dams ($d = N_f/N_m$), and each dam produced a total of $n_o$ offspring ($\frac{1}{2}n_o$ of each sex), so that the total number of offspring in a cohort equalled $T = n_o N_f$. Before reproductive age, the phenotype of individuals was recorded and a selection index was calculated. Because index weights were constant over time and no additional phenotypic information was included in the index at later ages, the index of individuals remained constant over time and the ranking of animals within categories remained unchanged over time. Within categories, individuals were ranked on the index, and the highest ranking $n_k$ individuals were selected. The number of parents selected from each category was determined in advance and remained constant over time, as in conventional gene flow (Hill 1974). Selection on estimated breeding value across categories, which gives the highest genetic level of the offspring in the next generation (James 1987), was not applied. The selection index was

$$I = b_1(P - \bar{P}_{FS}) + b_2(\bar{P}_{FS} - \bar{P}_{HS}) + b_3\bar{P}_{HS},$$

where $P$ is the phenotype of the individual, $\bar{P}_{FS}$ is the mean of $n_o$ full-sib records (including the individual), and $\bar{P}_{HS}$ is the mean of $n_o d$ half-sib records (including the individual and its full sibs). This form was used by Wray *et al.* (1994) and is convenient because the three sources of information are independent, which simplifies expressions such as the accuracy of selection. Note that mass selection is a special case of this index, where $b_1 = b_2 = b_3$. Different sets of index weights were chosen to allow for different selection strategies, *i.e.*, for a varying emphasis on family information.

**Basic approach for prediction of long-term genetic contributions:** This section introduces the concept of long-term genetic contributions. The long-term genetic contribution ($r_i$) of ancestor $i$ in cohort $t_1$ is defined as the proportion of genes present in all individuals in cohort $t_2$ deriving by descent from $i$, where $(t_2 - t_1) \rightarrow \infty$ (Woolliams *et al.* 1993). In other words, the long-term genetic contribution of an ancestor is the ultimate proportional contribution of the ancestor to generations in the distant future. After several generations, genetic contributions of ancestors stabilize (long-term contributions are reached) and become equal for all individuals in that and subsequent generations of descendants, but values differ between ancestors (Wray and Thompson 1990).

In the remainder of the current article, long-term genetic contributions of ancestors are referred to as "genetic contributions," unless explicitly stated otherwise. Applying the approach adopted by Woolliams *et al.* (1999), contributions of ancestors are predicted by conditioning on the selective advantage of those ancestors. Since sib indices are used here, the selective advantage is equal to the true breeding value of the ancestor [the only parental effect affecting selection of the offspring is the breeding value of the parent (Wray *et al.* 1994)]. For an individual in category $l$, $E(r_{i(l)}|A_{i(l)}) \approx u_{i(l)} = \alpha_l + \beta_l (A_{i(l)} - \bar{A}_l)$, where $\alpha_l$ is the expected contribution of an average parent in category $l$, $\beta_l$ is the

regression of the contribution of $i$ on its breeding value ($A_{i(l)}$), and $\bar{A}_l$ is the mean breeding value of selected contemporaries of $i$ in category $l$. For discrete generations, the complication of categories can be ignored and $\alpha$ is obtained directly from the number of parents: $\alpha = (2N_x)^{-1}$, ($x = m, f$; Wray and Thompson 1990). For both discrete and overlapping generations, solutions for $\beta$ can be obtained from two regression models (Woolliams 1998; Woolliams *et al.* 1999): first, the regression of the number of selected offspring on the breeding value of the parent ($\lambda$), and second, the regression of the breeding value of *selected* offspring on the breeding value of the parent ($\pi$). Both $\lambda$ and $\pi$ can be computed on the basis of known parameters; a derivation is in appendix a. Under equilibrium genetic parameters (Bulmer 1971), regression coefficients ($\alpha$, $\beta$, $\lambda$, $\pi$) are equal for the parental and offspring generation, allowing for the following closed form expression to compute $\beta$ instead of a recursive algorithm (Woolliams 1998):

$$\beta = (1 - \pi)^{-1}\lambda\alpha.$$

**Prediction of expected long-term genetic contributions in populations with overlapping generations:** This section describes the approach of Woolliams *et al.* (1999) to predict long-term genetic contributions for populations with overlapping generations. For ancestor $i$ in category $l$, the expected long-term genetic contribution was predicted from $u_{i(l)} = \alpha_l + \beta_l(A_{i(l)} - \bar{A}_l)$. Predictions of genetic contributions are obtained using a modified gene flow matrix ($\mathbf{G}$) of dimension $2c_{\max} \times 2c_{\max}$, which identifies the origin of genes of selected instead of newborn offspring. If the conventional gene flow matrix (Hill 1974) is denoted by $\mathbf{G}_0$, elements $g^0_{kl}$ represent the proportion of genes currently in category $k$ that were in category $l$ one time unit ago. In the modified gene flow matrix, elements $g_{kl}$ of $\mathbf{G}$ represent the proportion of genes in the $n_k$ selected individuals in category $k$ that were contributed by parents in category $l$. (Contributed by a parent in category $l$ refers to contribution via offspring that were *born* when the parent was in category $l$.) Because $\mathbf{G}$ represents the parental origin of the genes of selected individuals, it is affected by the degree of selection that is taking place, and this may vary with age. Because selected individuals may be born $c_{\max}$ years ago (and the age of parents *at birth* of offspring is relevant), $\mathbf{G}$ has a memory of $c_{\max}$ years, whereas $\mathbf{G}_0$ has only 1 year memory.

Solutions for $\alpha$ and $\beta$ were obtained from the basic equations (Woolliams *et al.* 1999),

$$\alpha_l = \sum_{k=1}^{2c_{\max}} n_l^{-1} g_{kl} n_k \alpha_k \quad (1)$$

$$\beta_l(A_{i(l)} - \bar{A}_l) = \sum_{k=1}^{2c_{\max}} n_l^{-1} g_{kl} n_k \alpha_k \lambda_{kl} (A_{i(k)} - \bar{A}_k) + \sum_{k=1}^{2c_{\max}} n_l^{-1} g_{kl} n_k \beta_k \pi_{kl} (A_{i(k)} - \bar{A}_k), \quad (2)$$

where $\lambda_{kl}$ is the regression coefficient of the selected number of offspring in category $k$ on the breeding value of the parent in category $l$, and $\pi_{kl}$ is the regression of the breeding value of selected offspring in category $k$ on the breeding value of the parent in category $l$. An intuitive understanding of Equations 1 and 2 can be gained by noting that $2n_l^{-1}g_{kl}n_k$ represents the average number of selected offspring in category $k$ of an ancestor in category $l$. Therefore, in (1), $\alpha_l$ is equal to $\frac{1}{2}$ times the sum of the average contributions of all selected offspring. (The other $\frac{1}{2}$ originates from the other parent.) In (1) it is implicitly assumed that the contribution of an average selected offspring ($\alpha_k$) is not dependent on the category of the parent ($l$). In (2), the first summation represents the change of contributions due to deviations of the selected number of offspring from the average. The second summation represents changes of genetic contributions of ancestors due to deviations of the breeding value of selected offspring of this ancestor from the average breeding value of selected contemporaries. In matrix form, combining Equations 1 and 2 for all categories $l$ (Woolliams *et al.* 1999),

$$\mathbf{N}\alpha = \mathbf{G}^{\mathrm{T}}\mathbf{N}\alpha \quad (3)$$

$$\mathbf{N}\beta = (\mathbf{I} - \mathbf{G}^{\mathrm{T}} * \Pi^{\mathrm{T}})^{-1}(\mathbf{G}^{\mathrm{T}} * \Lambda^{\mathrm{T}})(\mathbf{N}\alpha), \quad (4)$$

where * denotes the element-by-element multiplication, T denotes the transpose of matrices, $\mathbf{I}$ is a $2c_{\max} \times 2c_{\max}$ identity matrix, $\mathbf{N}$ is a $2c_{\max} \times 2c_{\max}$ diagonal matrix of elements $n_k$, $\Pi$ is a $2c_{\max} \times 2c_{\max}$ matrix of elements $\pi_{kl}$, $\Lambda$ is a $2c_{\max} \times 2c_{\max}$ matrix of elements $\lambda_{kl}$, $\alpha$ is a $2c_{\max}$ vector of elements $\alpha_l$, and $\beta$ is a $2c_{\max}$ vector of elements $\beta_l$. Throughout the article, matrices follow the gene flow notation, *i.e.*, rows represent offspring categories and columns represent parental categories. Prediction of genetic contributions using Equations 3 and 4 is referred to as Method M in results.

*Improved modified gene flow:* A first-order correction to Equation 1 was derived by taking account of differences among average breeding values of parental subgroups present in the selected offspring (Woolliams *et al.* 1999). When newborn offspring are grouped according to the category of parents, mean breeding values may differ between those groups. Selection then favors offspring descending from parental categories with a higher breeding value, increasing the genetic contribution of these categories. This phenomenon is fully accounted for by the modified gene flow matrix $\mathbf{G}$, identifying the origin of selected offspring. However, after selection, mean breeding values of selected offspring may still differ between parental category subgroups. This affects the contribution of categories, which was ignored in Equation 1. Improved prediction equations were obtained by conditioning on the parental category in Equation 1 (Woolliams *et al.* 1999),

$$\alpha_l = \sum_k n_l^{-1}\, g_{kl} n_k \{\alpha_k + \beta_k E[(A_{(i)k} - \bar{A}_k) \text{ given } i \text{ has category } l \text{ parent}]\}, \quad (5)$$

where $E[(A_{(i)k} - \bar{A}_k)$ given $i$ has category $l$ parent] is the expected breeding value of a selected offspring in category $k$ descending from a category $l$ parent, as deviation from the mean of selected contemporaries in category $k$. Substituting Equation 2 for β, the resulting expression is (Woolliams *et al.* 1999),

$$\mathbf{N}\alpha = [\mathbf{G}^{\mathrm{T}} + (\mathbf{G}^{\mathrm{T}} * \mathbf{D}^{\mathrm{T}})(\mathbf{I} - \mathbf{G}^{\mathrm{T}} * \Pi^{\mathrm{T}})^{-1}\,(\mathbf{G}^{\mathrm{T}} * \Lambda^{\mathrm{T}})]\mathbf{N}\alpha, \quad (6)$$

where $\mathbf{D}$ is a $2c_{\max} \times 2c_{\max}$ matrix of elements $d_{kl} = E[(A_{(i)k} - \bar{A}_k)$ given $i$ has category $l$ parent]. Therefore, Nα is obtained as a right eigenvector of the $2c_{\max} \times 2c_{\max}$ matrix $[\mathbf{G}^{\mathrm{T}} + (\mathbf{G}^{\mathrm{T}} * \mathbf{D}^{\mathrm{T}})(\mathbf{I} - \mathbf{G}^{\mathrm{T}} * \Pi^{\mathrm{T}})^{-1}\,(\mathbf{G}^{\mathrm{T}} * \Lambda^{\mathrm{T}})]$ with an eigenvalue of one (Woolliams *et al.* 1999). Solutions for β are still obtained from (4). Predictions of genetic contributions using Equations 4 and 6 will be referred to as Method P in results.

**Generation interval:** Generation interval ($L$) is defined as the turnover time of genes, *i.e.*, the average time interval between two meioses in which an average gene in the population is involved. This interval is equal to the time in which long-term genetic contributions sum to unity, *i.e.*, the genetic contribution summed over all ancestors entering the population over a time period of $L$ years equals unity: $\sum_L u_i = 1$. The generation interval (in years) is therefore equal to the reciprocal of the total long-term genetic contribution per year, *i.e.*, summed over all ancestors per year. In $u_{i(l)} = \alpha_l + \beta_l(A_{i(l)} - \bar{A}_l)$, the term $\beta(A_{i(k)} - \bar{A}_k)$ is zero on average, the sum of genetic contributions is therefore equal to $\sum_{k=1}^{2c\max} n_k\alpha_k$, and generation interval was calculated as (Woolliams *et al.* 1999),

$$L = 1 \Big/ \sum_{k=1}^{2c\max} n_k\alpha_k. \quad (7)$$

Generation intervals from this definition were compared to generation intervals defined as the average age of parents at birth of their offspring.

**Deterministic prediction procedure:** Elements of Equations 3 through 7 were obtained using an iterative procedure, which is described in this section. The iterative procedure is needed because elements (*e.g.*, variances, genetic gain, and genetic contributions) are mutually dependent and Bulmer's (1971) equilibrium parameters can only be reached by iteration. [Predictions can also be obtained using base generation parameters, but more accurate predictions are obtained using equilibrium parameters (Woolliams *et al.* 1999).] Predictions of genetic contributions shown in results are based on Bulmer's (1971) equilibrium parameters. A numerical example is in appendix c.

Phenotypic variance in year $t$ was the sum of additive genetic variance and environmental variance, $\sigma^2_{p,t} = \sigma^2_{A,t} + \sigma^2_E$. Environmental variance was constant over time. Additive genetic variance in an unselected cohort born at year $t$ was calculated as

$$\sigma^2_{A,t} = \sigma^2_{A(\mathrm{m}),t} + \sigma^2_{A(\mathrm{f}),t} + \tfrac{1}{2}\sigma^2_{A0},$$

where $\sigma^2_{A(\mathrm{m}),t}$ and $\sigma^2_{A(\mathrm{f}),t}$ are the between-sire and between-dam family additive genetic variance in unselected newborn offspring, and $\sigma^2_{A0}$ is the base generation additive genetic variance. Because genetic contributions are mainly determined in the first few generations, they are hardly affected by the rate of inbreeding. Therefore, no effect of inbreeding on the within-family variance was modeled.

Between-sire family additive genetic variance was calculated from

$$\sigma^2_{A(\mathrm{m}),t} = \tfrac{1}{4}\sum_{l=1}^{c\max} 2g^0_{1l}[\sigma^2_{A,t-1}(1 - \kappa_l\rho^2_{t-1}) + (\mu_{l,t-1} - \mu_{(\mathrm{m}),t-1})^2],$$

where $2g^0_{1l}$ is the proportion of offspring descending from sires in category $l$ ($2g^0_{1l} = n_l/N_{\mathrm{m}}$), $\kappa_l$ is Pearson's (1903) variance reduction coefficient, $\rho_t$ is the accuracy of the index in year $t$ (Wray *et al.* 1994), $\mu_{l,t}$ is the average breeding value of selected sires in category $l$, and $\mu_{(\mathrm{m}),t}$ is the average breeding value of all selected sires, *i.e.*, $\mu_{(\mathrm{m}),t} = \sum_{l=1}^{c\max} 2g^0_{1l}\mu_{l,t}$. Between-dam family additive genetic variance was calculated in the same way. For the calculation of $(\mu_{l,t-1} - \mu_{(\mathrm{m}),t-1})^2$, only differences between breeding values of selected individuals are important, and breeding values can be expressed relative to an arbitrary base. The genetic level of unselected animals at birth was taken as base here, and therefore, $\mu_{l,t} = (i_l\rho_t\sigma_{A,t} - \mathrm{age}(l)\Delta G_t)$, where $i_l$ is the selection intensity in category $l$ (not distinguishing between subgroups within categories and ignoring deviations from normality), and $\Delta G_t$ is the rate of genetic gain in year $t$. (It is assumed here that the difference between consecutive age classes is equal to $\Delta G$ from the last iteration, because this assumption decreases the number of iterations needed to reach equilibrium values that are not affected by the assumption.)

To calculate elements of the modified gene flow matrix, we need to find how the predefined selected proportion of individuals in category $k$ ($p_k$) is distributed across the parental age subgroups. The $k$th row of $\mathbf{G}$, therefore, was obtained by finding a common index truncation point for all parental subgroups represented among the selection candidates in category $k$ (separate for male and female parents). The solution for the common truncation point has to satisfy the equations (omitting subscript $t$ for simplicity)

$$p_k = \sum_l 2g^0_{1l}p_{kl}$$

$$p_{kl} = 1 - \Phi\left(\frac{I_k - \tfrac{1}{2}\tau_{(x)}\mu_l}{\sigma_{I,l}}\right),$$

where $p_{kl}$ is the selected proportion in the subclass de-

scending from parents in category $l$, $I_k$ is the index truncation point common for all offspring in category $k$, $\sigma_{I,l}$ is the standard deviation of the selection index of individuals descending from parents in category $l$, $\Phi$ denotes the cumulative normal density, and $\tau_{(x)}$ is twice the regression of the index of the offspring on the breeding value of the parent of sex $x$ ($x$ = m, f; Wray *et al.* 1994); *i.e.*, the term $\frac{1}{2}\tau_{(x)}\mu_{(x)l}$ represents the average index value of offspring descending from parents in category $l$. A solution for the common truncation point was obtained using the algorithm RIDDR_ROOT from *Numerical Recipes* (Press *et al.* 1992). Elements of **G** were derived from

$$g_{kl} = p_{kl}g_{1l}^{0}p_k^{-1}.$$

Elements of **D** are $d_{kl} = E[(A_{(i)k} - \bar{A}_k)$ given $i$ has category $l$ parent] and were calculated as (omitting subscript $t$ for simplicity)

$$d_{kl} = \tfrac{1}{2}\mu_l + i_{kl}\rho\sigma_{A_l} - \bar{d}_k,$$

with $\bar{d}_k = \sum_l 2g_{kl}[\frac{1}{2}\mu_l + i_{kl}\rho\sigma_{A_l}]$ calculated separately for each sex; where $i_{kl}$ is the selection intensity in subclass $kl$, and $\sigma_{A_l}$ is the additive genetic variance among unselected offspring descending from parents in category $l$.

Elements of $\Pi$ were calculated as $\pi_{kl} = \frac{1}{2}(1 - \kappa_k\tau_{(x)}\rho\sigma_A \times \sigma_I^{-1})$. Elements of $\Lambda$ were calculated as $\lambda_{kl} = \frac{1}{2}i_k\tau_{(x)}\sigma_I^{-1}$ (see appendix a). A general procedure to derive $\Pi$ and $\Lambda$ is in Woolliams *et al.* (1999).

As described in the section on prediction of long-term genetic contributions, $\boldsymbol{\alpha}$ can be obtained as a right eigenvector from Equation 3 for the "modified gene flow" and from Equation 6 for the "improved modified gene flow." In general, eigenvectors can be scaled, *i.e.*, if $\mathbf{x}$ is an eigenvector of matrix $\mathbf{A}$ with an eigenvalue $\gamma$, then $n\mathbf{x}$ will also be an eigenvector of $\mathbf{A}$ with the same eigenvalue $\gamma$. With the same eigenvalue, therefore, different eigenvectors can be obtained from Equations 3 or 6, and an additional constraint has to be imposed. Because contributions have to sum to unity per generation, the eigenvector was scaled accordingly. Therefore, first generation interval was calculated as the average age at birth of offspring weighted by the long-term genetic contribution of the categories ($n_k\alpha_k$): $L = \sum_{k=1}^{2c_{\max}} k n_k\alpha_k / \sum_{k=1}^{2c_{\max}} n_k\alpha_k$. And second, $\alpha$ was scaled so that $\sum_{k=1}^{2c_{\max}} n_k\alpha_k = L^{-1}$, *i.e.*, $\alpha$ is defined per year, and by definition the generation interval is the time in which contributions sum to unity.

Using $E(\Delta G) = \sum_{k=1}^{2c_{\max}} n_k E[r_{i(k)}a_{i(k)}]$, where $a_{i(k)}$ is the Mendelian sampling value of $i$ (Woolliams and Thompson 1994), genetic gain was predicted from $E(\Delta G_t) = \frac{1}{2}\sigma_{A_0}^2[\tau_w\sigma_I^{-1} \times \sum_{k=1}^{2c_{\max}} n_k\alpha_k i_k + \sum_{k=1}^{2c_{\max}} n_k\beta_k(1 - \kappa_k\tau_w\rho\sigma_A\sigma_I^{-1})]_t$, where $\tau_w$ is the regression of the index on the Mendelian sampling effect of the individual. A derivation is in appendix b.

**Stochastic simulation:** To draw inferences on the accuracy of predicted genetic contributions, the breeding scheme described in the *Population model* section was simulated stochastically and genetic contributions were estimated from simulated data. A noninbred and unselected base population of the appropriate family structure was generated. Breeding values of base population animals were taken from $N(0, \sigma_{A0}^2)$, and environmental values were from $N(0, \sigma_E^2)$. Within categories, individuals were ranked on the index, and the highest ranking $n_k$ individuals were selected from the $k$th category. Breeding values of offspring were obtained as $\frac{1}{2}A_m + \frac{1}{2}A_f + a$, where $A_m$, $A_f$, and $a$ are the sire and dam breeding values and the Mendelian sampling value. No effect of inbreeding on the Mendelian sampling variance was simulated, *i.e.*, $a \sim N(0, \frac{1}{2}\sigma_{A0}^2)$.

For the calculation of genetic contributions, an ancestor cohort $t_1$ was chosen when Bulmer's (1971) equilibrium genetic parameters were reached. Repeated cycles of selection and random mating were performed until genetic contributions were converged and a descendant cohort $t_2$ was chosen. Convergence time of genetic contributions ($t_2 - t_1$) was approximately equal to $7c_{\max}$. The long-term genetic contribution of ancestor $i$ in category $l$ in cohort $t_1$ to individuals in cohort $t_2$ was obtained by summing contributions via all pedigree paths leading from $i$ to individuals in $t_2$, $r_{i(l)} = T^{-1}\sum_{j=1}^{T} r_{i(l),j}$, where $r_{i(l),j}$ is the contribution to individual $j$ in cohort $t_2$. $r_{i(l),j}$ was calculated as $\sum_{\text{paths}} \frac{1}{2}^{\check{n}-1}$, where $\check{n}$ is the total number of animals (including $i$ and $j$) in a pedigree path from $i$ to $j$.

Genetic contributions were analyzed using the model $r_{i(l)} = \alpha_l + \beta_l(A_{i(l)} - \bar{A}_l) + e_{i(l)}$. $\alpha$ was estimated as $\hat{\alpha}_l = n_l^{-1}\sum_{i=1}^{n_l} r_{i(l)}$ and $\beta$ was estimated as $\hat{\beta}_l = \sum_{i=1}^{n_l} r_{i(l)}(A_{i(l)} - \bar{A}_l) / \sum_{i=1}^{n_l}(A_{i(l)} - \bar{A}_l)^2$. Asymptotic rate of genetic gain was calculated as $\Delta G = (G_{t2} - G_{t1})/(t_2 - t_1)$, where $G_t$ is the average breeding value of all animals born in cohort $t$. Generation interval was calculated as $L = 1/\sum_{k=1}^{2c_{\max}} n_k\hat{\alpha}_k$. Results were averaged over 500 replicates and standard errors were calculated from the variance among replicates.

## RESULTS

In this section, a comparison is made between results from conventional gene flow (Method C; Hill 1974), simple modified gene flow (Method M, Equations 3 and 4), and improved modified gene flow (Method P, Equations 4 and 6), for mass and sib-index selection.

### Mass selection

**Accuracy of $\alpha$:** Table 1 shows long-term genetic contributions of categories ($n_k\alpha_k$) obtained from conventional gene flow (Hill 1974) from Method C, Method M, Method P, and from simulation, for a population with three age classes, with 20 sires in age class 1, 10 dams in age class 1, and 30 dams in age class 3, *i.e.*, $\mathbf{N}$ = diag{20,0,0,10,0,30}. This scheme, with a high proportion of dams selected from the oldest age class, was

**TABLE 1**

**Genetic contributions of categories ($n_k\alpha_k$) under mass selection**

| $h_0^{2c}$ | Method C[a] | | | Method M[a] | | | Method P[a] | | | Simulation[b] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n_1\alpha_1$ | $n_4\alpha_4$ | $n_6\alpha_6$ | $n_1\alpha_1$ | $n_4\alpha_4$ | $n_6\alpha_6$ | $n_1\alpha_1$ | $n_4\alpha_4$ | $n_6\alpha_6$ | $n_1\alpha_1$ | $n_4\alpha_4$ | $n_6\alpha_6$ |
| 0.01 | 0.2857 | 0.0714 | 0.2143 | 0.2863 | 0.0726 | 0.2137 | 0.2869 | 0.0738 | 0.2131 | 0.2871 | 0.0758 | 0.2135 |
| | −0.0014 | −0.0044 | +0.0008 | −0.0008 | −0.0032 | +0.0002 | −0.0002 | −0.0020 | −0.0004 | 0.0008 | 0.0012 | 0.0012 |
| 0.20 | 0.2857 | 0.0714 | 0.2143 | 0.2979 | 0.0957 | 0.2021 | 0.3097 | 0.1194 | 0.1903 | 0.3107 | 0.1172 | 0.1924 |
| | −0.0250 | −0.0458 | +0.0219 | −0.0128 | −0.0215 | +0.0097 | −0.0010 | +0.0022 | −0.0021 | 0.0013 | 0.0021 | 0.0021 |
| 0.40 | 0.2857 | 0.0714 | 0.2143 | 0.3122 | 0.1244 | 0.1878 | 0.3363 | 0.1726 | 0.1637 | 0.3340 | 0.1663 | 0.1640 |
| | −0.0483 | −0.0949 | +0.0503 | −0.0218 | −0.0419 | +0.0238 | +0.0023 | +0.0063 | −0.0003 | 0.0016 | 0.0025 | 0.0026 |
| 0.60 | 0.2857 | 0.0714 | 0.2143 | 0.3314 | 0.1629 | 0.1686 | 0.3666 | 0.2332 | 0.1334 | 0.3579 | 0.2200 | 0.1449 |
| | −0.0722 | −0.1485 | +0.0694 | −0.0265 | −0.0571 | +0.0237 | +0.0087 | +0.0133 | −0.0115 | 0.0018 | 0.0030 | 0.0027 |
| 0.80 | 0.2857 | 0.0714 | 0.2143 | 0.3615 | 0.2230 | 0.1385 | 0.3996 | 0.2992 | 0.1004 | 0.3893 | 0.2766 | 0.1104 |
| | −0.1036 | −0.2052 | +0.1039 | −0.0278 | −0.0536 | +0.0281 | +0.0103 | +0.0226 | −0.0100 | 0.0018 | 0.0031 | 0.0026 |
| 0.99 | 0.2857 | 0.0714 | 0.2143 | 0.4115 | 0.3229 | 0.0885 | 0.4324 | 0.3648 | 0.0676 | 0.4188 | 0.3381 | 0.0818 |
| | −0.1331 | −0.2667 | +0.1325 | −0.0073 | −0.0152 | +0.0067 | +0.0136 | +0.0267 | −0.0142 | 0.0017 | 0.0027 | 0.0024 |

For $\mathbf{N}$ = diag{20,0,0,10,0,30}, eight tested offspring per dam ($n_o$ = 8), ancestor cohort ($t_1$) = 10, and descendant cohort ($t_2$) = 35.

[a] Deviations from simulation results are on every second line.

[b] Standard errors are on every second line.

[c] $h_0^2$ denotes heritability.

chosen because it clearly illustrates the effect of selective advantage on contributions of categories.

Results from Method C are independent of heritability ($h_0^2$), but results from Method M, Method P, and from simulation are not. For $h_0^2 = 0.01$, results from all methods are practically identical because heritable effects play a minor role in that case. For higher heritabilities, Method C shows considerable overestimates of contributions from 3-yr-old dams ($n_6\alpha_6$), whereas Methods M and P are significantly closer, and from these, Method P is most accurate. For high heritabilities (>0.6), absolute differences between Method P and simulation are roughly only 10% of the errors from Method C, and for this particular scheme in the opposite direction. The large differences between Method C and simulation are partly caused by the distribution of parents across age classes in Table 1. Because most dams are selected from the oldest category, offspring from these dams will have a low breeding value, which will reduce their genetic contribution. When parents are selected across age classes, differences between Method C and simulation will be much smaller (see discussion).

Comparing Methods M and P to simulation results shows that the first-order correction improves the accuracy of the predicted long-term genetic contributions. In Equation 3, differences between selective advantage of selected offspring from different parental categories ($d_{kl}$) are ignored, resulting in underprediction of contributions of young categories and in overprediction of contributions of older categories (except for $h_0^2$ = 0.99, probably due to deviations from normality for this extreme case, which is of little practical importance).

**Accuracy of β:** Table 2 shows the regression coefficients of contributions on breeding values (β), from Method M, Method P, and from simulation, for $\mathbf{N}$ = diag{20,0,0,10,0,30}. Most predictions from Method P are within three times the standard error of simulation results, and the trends in predictions agree well with simulation results. Method P was slightly more accurate than Method M, particularly when modeling the differences between 1- and 3-yr-old females, *i.e.*, $\beta_4$ and $\beta_6$. In Method C, the effect of selective advantage is not modeled, *i.e.*, β is implicitly zero.

**Accuracy of genetic gain and generation interval:** Table 3 shows genetic gain per year and generation interval from Method C, Method M, Method P, and from simulation, for $\mathbf{N}$ = diag{20,0,0,10,0,30}. Generation interval was calculated from Equation 7. For Method C, generation interval from Equation 7 is identical to the average age of parents when their progeny are born and is obtained from $\mathbf{G}_0$. Generation intervals based on the average age of parents of selected offspring, as suggested by Bichard *et al.* (1973), are obtained from $\mathbf{G}$ (see example in appendix c) and are also in Table 3. Method C does not account for the effect of selection on genetic contributions and therefore results in higher generation

**TABLE 2**

**Regression coefficients of long-term genetic contributions on breeding values ($\beta_i$) under mass selection**

| | Method M[a] | | | Method P[a] | | | Simulation[b] | | |
|---|---|---|---|---|---|---|---|---|---|
| $h_0^{2c}$ | $\beta_1$ | $\beta_4$ | $\beta_6$ | $\beta_1$ | $\beta_4$ | $\beta_6$ | $\beta_1$ | $\beta_4$ | $\beta_6$ |
| 0.01 | 0.0228 | 0.0116 | 0.0114 | 0.0229 | 0.0116 | 0.0114 | 0.0233 | 0.0102 | 0.0128 |
| | −0.0005 | +0.0014 | −0.0014 | −0.0004 | +0.0014 | −0.0014 | 0.0013 | 0.0017 | 0.0008 |
| 0.20 | 0.0211 | 0.0136 | 0.0095 | 0.0221 | 0.0144 | 0.0100 | 0.0205 | 0.0159 | 0.0093 |
| | +0.0006 | −0.0023 | +0.0002 | +0.0016 | −0.0015 | +0.0007 | 0.0005 | 0.0008 | 0.0003 |
| 0.40 | 0.0201 | 0.0161 | 0.0081 | 0.0220 | 0.0182 | 0.0086 | 0.0217 | 0.0205 | 0.0078 |
| | −0.0016 | −0.0044 | +0.0003 | +0.0003 | −0.0023 | +0.0008 | 0.0005 | 0.0009 | 0.0002 |
| 0.60 | 0.0199 | 0.0197 | 0.0067 | 0.0222 | 0.0233 | 0.0070 | 0.0199 | 0.0246 | 0.0065 |
| | +0.0000 | −0.0049 | +0.0002 | +0.0023 | −0.0013 | +0.0005 | 0.0006 | 0.0010 | 0.0002 |
| 0.80 | 0.0204 | 0.0254 | 0.0052 | 0.0226 | 0.0298 | 0.0051 | 0.0210 | 0.0291 | 0.0053 |
| | −0.0006 | −0.0037 | −0.0001 | +0.0016 | +0.0007 | −0.0002 | 0.0007 | 0.0012 | 0.0002 |
| 0.99 | 0.0220 | 0.0347 | 0.0031 | 0.0230 | 0.0372 | 0.0029 | 0.0246 | 0.0362 | 0.0047 |
| | −0.0026 | −0.0015 | −0.0016 | −0.0016 | +0.0010 | −0.0018 | 0.0012 | 0.0021 | 0.0003 |

For $\mathbf{N}$ = diag{20,0,0,10,0,30}, eight tested offspring per dam ($n_0$ = 8), ancestor cohort ($t_1$) = 10, and descendant cohort ($t_2$) = 35.

[a] Deviations from simulation results are on every second line.

[b] Standard errors are on every second line.

[c] $h_0^2$ denotes heritability.

intervals than simulation. For the scheme in Table 3, most dams are selected from the oldest category, which increases differences between Method C and Method P. Even when the numbers of females selected were exchanged, however, *i.e.*, $\mathbf{N}$ = diag{20,0,0,30,0,10}, there were differences between generation intervals from Method C and Method P (see Figure 2). Method M showed systematic overprediction of generation intervals, which agrees with the overprediction of contributions of older categories (see Table 1). Predicted generation intervals based on the average age of parents of selected offspring, *i.e.*, from $\mathbf{G}$ rather than $\mathbf{G}_0$, were very close to generation intervals from Method M. Generation intervals from Method P were close to simulation results, only showing minor underprediction for high heritabilities.

For this particular scheme, genetic gain from Method C was more accurate than gain from Method P. However, this was not a general result; *e.g.*, for $\mathbf{N}$ = diag {20,0,0,30,0,10} (results not shown) it was the other way

**TABLE 3**

**Rate of genetic gain ($\Delta G$) and generation interval ($L$) under mass selection**

| | Method C[a] | | Method M[a] | | Method P[a] | | Selected offspring[a] | Simulation[b] | |
|---|---|---|---|---|---|---|---|---|---|
| $h_0^{2c}$ | $\Delta G$ | $L$ | $\Delta G$ | $L$ | $\Delta G$ | $L$ | $L$ | $\Delta G$ | $L$ |
| 0.01 | 0.0090 | 1.750 | 0.0091 | 1.746 | 0.0092 | 1.743 | 1.746 | 0.0089 | 1.735 |
| | +0.0001 | +0.015 | +0.0002 | +0.011 | +0.0002 | +0.008 | +0.011 | 0.0005 | 0.005 |
| 0.20 | 0.1751 | 1.750 | 0.1687 | 1.679 | 0.1770 | 1.615 | 1.678 | 0.1702 | 1.612 |
| | +0.0049 | +0.138 | −0.0015 | +0.067 | +0.0068 | +0.003 | +0.066 | 0.0005 | 0.007 |
| 0.40 | 0.3386 | 1.750 | 0.3222 | 1.602 | 0.3513 | 1.487 | 1.595 | 0.3363 | 1.505 |
| | +0.0023 | +0.245 | −0.0141 | +0.097 | +0.0150 | −0.018 | +0.090 | 0.0008 | 0.008 |
| 0.60 | 0.4960 | 1.750 | 0.4774 | 1.509 | 0.5325 | 1.364 | 1.500 | 0.5044 | 1.384 |
| | −0.0084 | +0.366 | −0.0270 | +0.125 | +0.0281 | −0.020 | +0.116 | 0.0009 | 0.008 |
| 0.80 | 0.6518 | 1.750 | 0.6533 | 1.383 | 0.7224 | 1.251 | 1.381 | 0.6749 | 1.288 |
| | −0.0231 | +0.462 | −0.0216 | +0.095 | +0.0475 | −0.037 | +0.093 | 0.0009 | 0.006 |
| 0.99 | 0.8024 | 1.750 | 0.8691 | 1.215 | 0.9098 | 1.156 | 1.244 | 0.8420 | 1.192 |
| | −0.0396 | +0.558 | +0.0271 | +0.023 | +0.0678 | −0.036 | +0.052 | 0.0009 | 0.005 |

For $\mathbf{N}$ = diag{20,0,0,10,0,30}, eight tested offspring per dam ($n_0$ = 8), ancestor cohort ($t_1$) = 10, and descendant cohort ($t_2$) = 35.

[a] Deviations from simulation results are on every second line.

[b] Standard errors are on every second line.

[c] $h_0^2$ denotes heritability.

around. In general, both methods showed similar accuracies for predicting genetic gain.

**Effect of heritability and selection intensity on α:** The effect of heritability and selection intensity on average genetic contributions of categories ($n_k\alpha_k$) was studied using Method P. Figure 1 shows the predicted long-term genetic contribution of 1-yr-old females as a proportion of the total contribution of females ($n_4\alpha_4/(n_4\alpha_4 + n_6\alpha_6)$), for two different breeding schemes and for two selection intensities. The breeding schemes were $S_1$: **N** = diag{20,0,0,30,0,10} and $S_2$: **N** = diag{20,0,0,10,0,30}. Selection intensity was varied by varying the number of tested offspring per dam, *i.e.*, $n_o$ was 4 or 20. To illustrate the relation between genetic contributions and generation interval, Figure 2 shows the corresponding generation interval. In $S_1$ and $S_2$, males are selected from a single age, and $L$ is directly related to $n_4\alpha_4/(n_4\alpha_4 + n_6\alpha_6)$. Results from Method C are identical to results for $h^2 = 0$.

Figure 1 clearly shows an increased contribution of 1-yr-old females when heritability increases, which is due to an increased selective advantage of offspring descending from 1-yr-old dams when $h_0^2$ increases. As heritability increased from 0.2 to 0.8, genetic gain per year increased from 0.232 to 0.977 units $\sigma_p$ for **N** = diag{20,0,0,10,0,30} and $n_0 = 20$. Consequently, the difference between average breeding values of offspring from 1- and 3-yr-old dams increased from 0.277 to 1.153. This selective advantage resulted in an increased proportion of offspring selected from 1-yr-old dams when $h_0^2$ increased. When $h_0^2$ increased from 0.2 to 0.8, it showed that among the 1-yr-old selected females, the proportion descending from 1-yr-old dams increased from 0.386 to 0.894. [These proportions were determined from the **G** matrix (not shown).]

The relative long-term genetic contribution of 1-yr-old females also increased with $n_o$ (see Figure 1), *i.e.*, with selection intensity. This is partly due to increased genetic gain resulting in an increased selective advantage of newborn offspring of 1-yr-old dams, in the same way as when $h_0^2$ increases, but also due to a decreased overall selected proportion moving the common truncation point of subclasses to the right. When a common truncation point for two normal distributions with different means is moved to the right, the smaller upper-tail probability of the two will decrease more rapidly than the larger upper-tail probability, due to the nonlinear relation between truncation point and selected proportion, therefore decreasing the relative contribution of 3-yr-olds. This effect can be illustrated by comparing the relative contribution of 1-yr-old females between schemes with different selection intensities at the same $\Delta G$, because with the same $\Delta G$ the difference between mean breeding values of 1- and 3-yr-old dams will be the same. For **N** = diag{20,0,0,10,0,30}, $n_o = 20$ and $h_0^2 = 0.4$, $\Delta G$ was 0.4854, and the same $\Delta G$ can be obtained with identical **N**, but with $n_o = 4$ and $h_0^2 = 0.77$. However, the relative contribution of 1-yr-old females differed considerably: 0.685 for $n_o = 20$ compared to 0.540 for $n_o = 4$ (see Figure 1), mainly due to different selection intensities.

**Effect of selection intensity on β:** Figure 3 shows the relation between selection intensity and β for a scheme

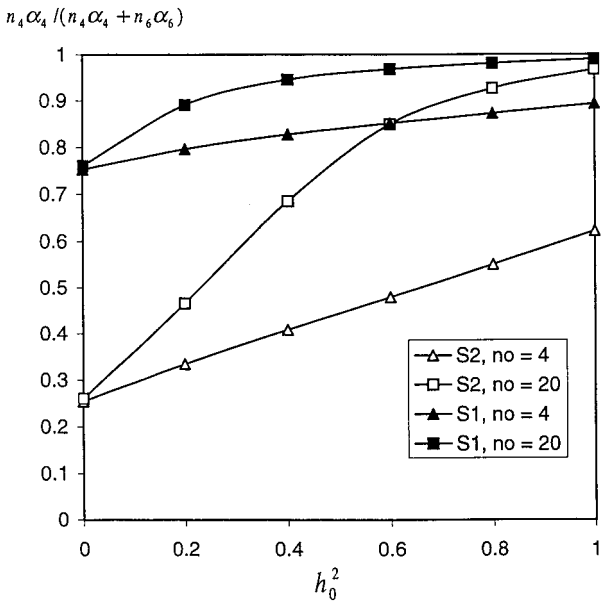


Figure 1.—Predicted long-term genetic contributions from Method P of 1-yr-old females as a proportion of the total contribution of females $n_4\alpha_4/(n_4\alpha_4 + n_6\alpha_6)$, as a function of heritability, for $S_1$: **N** = diag{20,0,0,30,0,10) and $S_2$: **N** = diag{20,0,0,10,0,30}, for $n_o = 4$ or $n_o = 20$ tested offspring per dam, and mass selection ($b_1 = b_2 = b_3$).

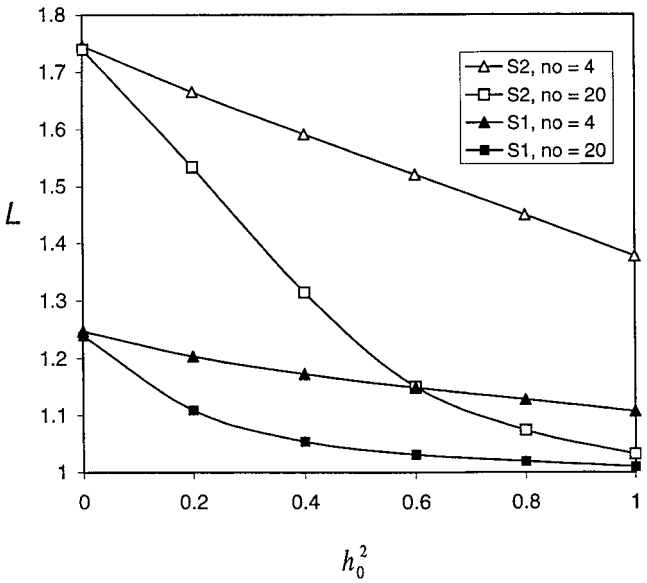


Figure 2.—Predicted generation interval ($L$) from Method P, as a function of heritability, for $S_1$: **N** = diag{20,0,0,30,0,10) and $S_2$: **N** = diag{20,0,0,10,0,30}, for $n_o = 4$ or $n_o = 20$ tested offspring per dam, and mass selection ($b_1 = b_2 = b_3$).

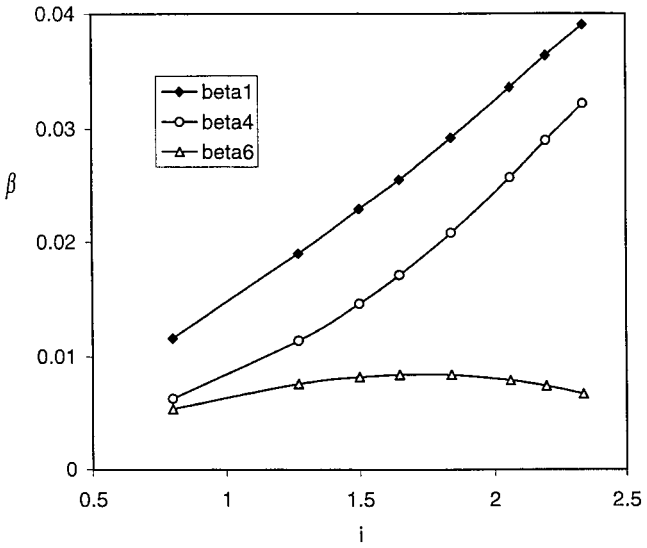


Figure 3.—Predicted regression coefficients of long-term genetic contributions on breeding values (β) from Method P as a function of selection intensity ($i$), for age class 1 males ($\beta_1$), age class 1 females ($\beta_4$), and age class 3 females ($\beta_6$), for $h_0^2$ = 0.4, **N** = diag{20,0,0,20,0,20}, and mass selection ($b_1 = b_2 = b_3$).

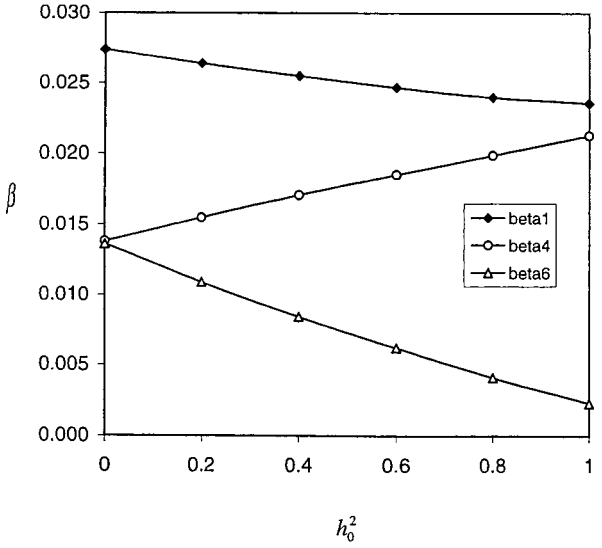


Figure 4.—Predicted regression coefficients of long-term genetic contributions on breeding values (β) from Method P as a function of heritability ($h_0^2$), for age class 1 males ($\beta_1$), age class 1 females ($\beta_4$), and age class 3 females ($\beta_6$), for **N** = diag{20,0,0,20,0,20}, $n_o$ = 8 tested offspring per dam, and mass selection ($b_1 = b_2 = b_3$).

with **N** = diag{20,0,0,20,0,20} using Method P. Selection intensity is equal for all categories in this scheme, and was varied by varying the number of tested offspring per dam from $n_o = 2$ ($i = 0.798$) to $n_o = 40$ ($i = 2.336$).

Figure 3 shows an increase in $\beta_1$ and $\beta_4$ with increasing selection intensity. On average, β is expected to increase with selection intensity because the regression of selected number of offspring on breeding value (λ) increases with selection intensity (see appendix a) and β is positively related to λ (see Equation 2), explaining the trend for $\beta_1$ and $\beta_4$. For $\beta_6$ the increase with selection intensity is counteracted by the reduced total contribution of 3-yr-old dams (see Figure 1). For other heritabilities (results not shown) the relation between β and selection intensity was similar.

**Effect of heritability on β:** Figure 4 shows the relation between β and heritability using Method P. For $h_0^2 = 0$, $\beta_4 = \beta_6 = \frac{1}{2}\beta_1$, which is to be expected from (2) when selection intensity is equal for all categories and $g_{kl} = g_{kl}^0$ because $h_0^2 = 0$. When $h_0^2$ increases, genetic gain increases, resulting in a higher proportion of selected offspring descending from 1-yr-old parents, *i.e.*, for all $k$, $g_{k4} > g_{k6}$ for $h_0^2 > 0$. When $g_{k4} > g_{k6}$ and selection intensity is equal for all categories, it can be inferred from (2) that $\beta_4 > \beta_6$ as in Figure 4.

It is a general conclusion for mass selection, therefore, that β of younger categories will increase with $h_0^2$, whereas β of older categories will decrease with $h_0^2$. The interpretation of this relation is, that under mass selection the contributions of young animals will increasingly be determined by their breeding value when $h_0^2$ increases, whereas for older animals the effect of breeding value on contributions will decrease with increasing $h_0^2$. An intuitive way of looking at this is, that for influential animals (which are young animals when $h_0^2$ is high) a change of breeding value gives a larger (absolute) change of genetic contributions than it does for unimportant animals. The same reasoning holds for the relation between β and selection intensity, explaining the different trend of $\beta_4$ and $\beta_6$ in Figure 3.

The regression coefficient for 1-yr-old males ($\beta_1$) shows only minor variation with $h_0^2$ because males are selected from a single category in Figure 3. Therefore, category 1 always contributes 50% of the genes of selected offspring ($g_{11} = g_{41} = g_{61} = 0.5$) regardless of heritability, and variation of $\beta_1$ with $h_0^2$ is only due to variation in λ and π.

## Selection on a sib index

Long-term genetic contributions of categories ($n_k\alpha_k$) are mainly dependent on the modified gene flow matrix. For a sib index, **G** is determined by genetic gain and selected proportions, in the same way as for mass selection. The main differences between sib index and mass selection are, therefore, in the regressions λ and π, resulting in different predictions for β. Results for a sib index, therefore, focus on β, though α will also differ from results for mass selection.

**Accuracy of β:** Predictions for a sib index are compared to simulation results for two opposite schemes: a scheme with positive weight on family information and a scheme with negative weight on family information. The weights used are different from the classical se-

**TABLE 4**

**Regression coefficients of long-term genetic contributions on breeding values (β) for a sib index with positive weight on family information**

| | Method P[a] | | | Simulation[b] | | |
|---|---|---|---|---|---|---|
| $h_0^{2c}$ | $\beta_1$ | $\beta_4$ | $\beta_6$ | $\beta_1$ | $\beta_4$ | $\beta_6$ |
| 0.01 | 0.0391 | 0.0188 | 0.0181 | 0.0359 | 0.0149 | 0.0175 |
| | +0.0032 | +0.0039 | +0.0006 | 0.0019 | 0.0021 | 0.0010 |
| 0.20 | 0.0346 | 0.0251 | 0.0132 | 0.0310 | 0.0255 | 0.0113 |
| | +0.0036 | −0.0004 | +0.0019 | 0.0008 | 0.0012 | 0.0004 |
| 0.40 | 0.0325 | 0.0321 | 0.0096 | 0.0307 | 0.0295 | 0.0081 |
| | +0.0018 | +0.0026 | +0.0015 | 0.0008 | 0.0013 | 0.0003 |
| 0.60 | 0.0311 | 0.0389 | 0.0064 | 0.0280 | 0.0339 | 0.0077 |
| | +0.0031 | +0.0050 | −0.0013 | 0.0008 | 0.0014 | 0.0003 |
| 0.80 | 0.0299 | 0.0448 | 0.0037 | 0.0295 | 0.0397 | 0.0053 |
| | +0.0004 | +0.0051 | −0.0016 | 0.0010 | 0.0017 | 0.0003 |
| 0.99 | 0.0293 | 0.0498 | 0.0044 | 0.0293 | 0.0404 | 0.0040 |
| | −0.0000 | +0.0094 | +0.0004 | 0.0011 | 0.0021 | 0.0004 |

For $\mathbf{N}$ = diag{20,0,0,10,0,30}, $n_o = 8$, $b_1 = 1.0$, $b_2 = 1.5$, $b_3 = 2$, ancestor cohort ($t_1$) = 10, and descendant cohort ($t_2$) = 35.

[a] Deviations from simulation results are on every second line.

[b] Standard errors are on every second line.

[c] $h_0^2$ denotes heritability.

lection index weights (Hazel 1943), but as shown by Villanueva and Woolliams (1997), optimum index weights for intermediate and long-term responses are generally different from classical index weights.

For positive weight on family information, Table 4 shows β from Method P and from simulation for $\mathbf{N}$ = diag{20,0,0,10,0,30}, $b_1 = 1$, $b_2 = 1.5$, and $b_3 = 2$ (*i.e.*, $I = P + \frac{1}{2}\bar{P}_{FS} + \frac{1}{2}\bar{P}_{HS}$). In Table 4, Method P shows the same trend as simulation results, but tends to slightly overestimate regression coefficients for 1-yr-old parents ($\beta_1$ and $\beta_4$). Predictions of α (results not shown) were close to simulation results and showed similar trends as for mass selection.

For negative weight on family information, Table 5 shows β from Method P and from simulation, for $\mathbf{N}$ = diag{20,0,0,10,0,30}, $b_1 = 1$, $b_2 = 0.5$, and $b_3 = 0$ (*i.e.*, $I = P - \frac{1}{2}\bar{P}_{FS} - \frac{1}{2}\bar{P}_{HS}$). In Table 5, Method P shows the same trend as simulation results and is accurate. Predictions for α (results not shown) were very accurate, *i.e.*, within ±3 SE with 500 replicates in the simulation.

**Effect of index weights on β:** Figure 5 shows the effect of a varying emphasis on family information in the selection index on the regression coefficients of long-term genetic contributions on breeding values, for 1-yr-old male parents ($\beta_1$), from Method P (lines), and from simulation (markers) for $\mathbf{N}$ = diag{20,0,0,20,0,20}. For this scheme, $\beta_1$ gives a good impression of the average level of β, because males are selected from a single category, *i.e.*, there is no competition between categories going on. In Figure 5, the index weights vary from $b_1 = 1$, $b_2 = b_3 = 0$, representing complete within-family selection, to $b_1 = 1$, $b_2 = 2$, $b_3 = 2$, which is identical to $I = P + \bar{P}_{FS}$.

For within-family selection, β equals zero because offspring are selected on their Mendelian sampling term, which by definition is independent of the parental breeding value. Therefore, selective advantage is not inherited and results (both α and β) are identical to results from Method C.

When index weights on family information increased, $\beta_1$ increased because selection of offspring is increasingly affected by the parental breeding value. Similar relations between the average level of β and weight given to family information were found for other distributions of parents across categories (including schemes with competition between categories).

When weight on family information increases, selection tends to selection of families instead of individuals, whereas λ is derived assuming a continuous linear change. Accuracy of predictions decreased, therefore, when weight given to family information became high, which is shown by the increased difference between lines and markers in Figure 5.

## DISCUSSION

This article has studied in detail two methods proposed by Woolliams *et al.* (1999) for the prediction of long-term genetic contributions of individuals in selected populations with overlapping generations. The methods enable accurate predictions of long-term genetic contributions of individual animals and of catego-

**TABLE 5**

**Regression coefficients of long-term genetic contributions on breeding values (β) for a sib index with negative weight on family information**

| | Method P[a] | | | Simulation[b] | | |
|---|---|---|---|---|---|---|
| $h_0^{2c}$ | $\beta_1$ | $\beta_4$ | $\beta_6$ | $\beta_1$ | $\beta_4$ | $\beta_6$ |
| 0.01 | 0.0015 | 0.0023 | 0.0023 | 0.0013 | 0.0007 | 0.0018 |
| | +0.0002 | +0.0016 | +0.0005 | 0.0011 | 0.0012 | 0.0006 |
| 0.20 | 0.0016 | 0.0025 | 0.0024 | 0.0016 | 0.0023 | 0.0023 |
| | +0.0000 | +0.0002 | +0.0001 | 0.0002 | 0.0003 | 0.0001 |
| 0.40 | 0.0017 | 0.0028 | 0.0024 | 0.0016 | 0.0030 | 0.0024 |
| | +0.0001 | −0.0002 | +0.0000 | 0.0002 | 0.0002 | 0.0001 |
| 0.60 | 0.0018 | 0.0032 | 0.0025 | 0.0016 | 0.0034 | 0.0025 |
| | +0.0002 | −0.0002 | +0.0000 | 0.0002 | 0.0002 | 0.0001 |
| 0.80 | 0.0019 | 0.0037 | 0.0026 | 0.0021 | 0.0045 | 0.0026 |
| | −0.0002 | −0.0008 | +0.0000 | 0.0002 | 0.0003 | 0.0001 |
| 0.99 | 0.0021 | 0.0045 | 0.0027 | 0.0024 | 0.0044 | 0.0026 |
| | −0.0003 | +0.0001 | +0.0001 | 0.0002 | 0.0003 | 0.0001 |

For **N** = diag{20,0,0,10,0,30}, eight tested offspring per dam ($n_o = 8$, $b_1 = 1.0$, $b_2 = 0.5$, $b_3 = 0$, ancestor cohort ($t_1$) = 10, and descendant cohort ($t_2$) = 35.

[a] Deviations from simulation results are on every second line.

[b] Standard errors are on every second line.

[c] $h_0^2$ denotes heritability.

ries using a simple linear model. Predictions of genetic contributions within categories were first shown by Woolliams *et al.* (1999) but never studied in detail. Genetic contributions were predicted conditional on breeding value and category of the ancestor by using a modified gene flow approach. The method accounts for the inheritance of selective advantage both between and within categories, resulting in more accurate predictions of genetic contributions and generation intervals than methods based on contributions to newborn offspring in the next cohort. Some trends in the prediction errors remained (*e.g.*, Table 1, Figure 5), but this is merely a matter of improving the relevant regression equations; they do not undermine the basic ideas underlying the theory. Conventional methods ignore the effect of selection on genetic contributions and therefore underestimate contributions of younger categories and overestimate generation interval. Thus, improved methods were necessary.

Accurate predictions of long-term genetic contributions for overlapping generation schemes facilitate deterministic prediction of rates of inbreeding for these schemes (Woolliams 1998) and consequently enable a computationally feasible optimization of breeding schemes with restricted inbreeding. The modified gene flow approach enables prediction of *individual* long-term genetic contributions [by including $\beta_k(A_{i(k)} - \bar{A}_k)$ in the model for expected contributions], whereas conventional gene flow only enables prediction of average genetic contributions (*i.e.*, assuming $\beta = 0$). For the prediction of rates of inbreeding it is crucial to account for the effect of selection between individuals (Wray *et al.* 1990), and conventional gene flow is therefore not suitable for prediction of rates of inbreeding.

In the present study, generation interval was defined as $L = 1/\Sigma n_k\alpha_k$, *i.e.*, the generation interval is the time in which long-term genetic contributions sum to unity. Intuitively, this is a sensible definition: One generation is the time in which the genes are turned over once. The definition of generation interval as the time in which contributions sum to unity is general and is also applicable to generation intervals based on newborn progeny or on selected progeny. For example, generation interval based on newborn progeny, *i.e.*, the average age of parents when progeny are born, can also be calculated as $L_0 = 1/\Sigma\alpha_0$, where $\alpha_0$ are contributions obtained from conventional gene flow. Generation interval based on contributions to selected offspring only ($L_1$), *i.e.*, the average age of parents of selected offspring, can be obtained from the modified gene flow matrix **G** (see appendix c) and was close to results from simple modified gene flow. When genetic gain is made and selective advantage is inherited, generation interval based on long-term genetic contributions is shorter than both $L_0$ and $L_1$, because selective advantage is partly passed on to more distant offspring.

Whereas $L_0$ and $L_1$ are based on contributions at a specific time point, *i.e.*, before and immediately after selection of the offspring, $L$ is based on converged, *i.e.*, asymptotic long-term genetic contributions of parental categories, which are an invariable property of a population once contributions have converged. Therefore, the definition of generation interval based on long-term genetic contributions is equal to the turnover time of

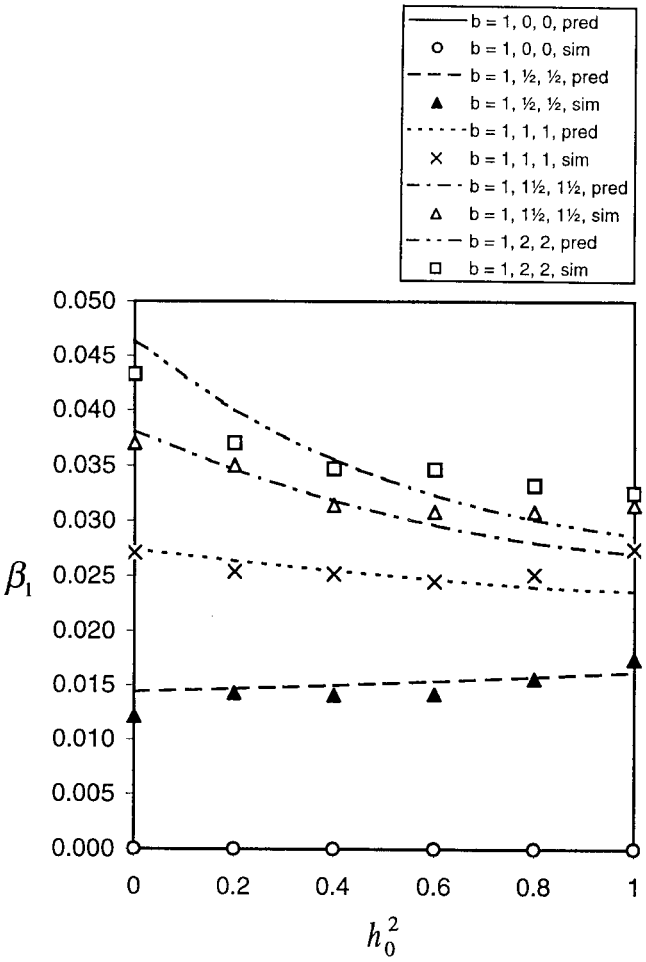


Figure 5.—Predicted and simulated regression coefficients of long-term genetic contributions on breeding values for 1-yr-old males ($\beta_1$) from Method P, for different index weighting factors ($b$) and a range of heritabilities ($h_0^2$). For $\mathbf{N}$ = diag{20,0,0,20,0,20} and $n_o$ = 8. Lines indicate predictions, markers indicate results from simulation.

genes, *i.e.*, it is the average time interval between two meioses, and it is of a more genetical and less operational nature than $L_0$ and $L_1$.

In the present study, results are only presented for situations where the selection index of an animal was constant across ages. In practice, animals in different categories will often have different amounts of information, affecting the variance of the selection index. This will mainly affect the **G** matrix, but is easily accounted for by using index variances specific to categories in the equations presented in methods. The problem is more complex for the prediction of rates of inbreeding, because in that case the lifetime genetic contribution of an ancestor, *i.e.*, its contribution summed over all categories it belonged to over its entire life, is relevant, which requires the probability that the same animal was selected in multiple categories.

Large differences were found between predicted genetic contributions from conventional and from modified gene flow in the present study. These differences were partly caused by the distribution of parents across categories; *i.e.*, in Tables 1 and 3 the majority of the dams were selected from the oldest category. When animals are selected by truncation across categories, differences in generation interval between the two methods will be much smaller. For example, for $h_0^2$ = 0.5, $n_o$ = 4, $N_m$ = 20, $N_f$ = 40, truncation selection across categories resulted in $\mathbf{N}$ = diag{18,2,33,7}, predicted generation interval from conventional gene flow was 1.138 and from modified gene flow was 1.129 (simulation: $L$ = 1.130). An advantage of modified gene flow is that it gives accurate predictions of generation interval for any arbitrary distribution of parents across categories, and it is not limited to truncation selection across categories.

In the present article, the within-family variance was assumed to be constant over time, which is not strictly true when inbreeding is accumulating. However, genetic contributions are mainly determined in the first few generations, where the inbreeding effects on descendants are still small. Long-term genetic contributions are therefore hardly affected by a reduction of variance due to inbreeding. Furthermore, ignoring the effect of inbreeding on the variance allows for the assumption of Bulmer's (1971) equilibrium variances (assuming the infinitesimal model), which greatly simplifies prediction equations for long-term genetic contributions (Woolliams *et al.* 1999). For extremely small populations, *e.g.*, with fewer than five parents per sex, it may become important to account for the effect of inbreeding when predicting long-term genetic contributions.

The number of parents is no guarantee for the genetic constitution of populations in the long term, because selective advantage of parents is inherited by offspring. This is a point of concern for conservation genetics where genetic improvement is also being sought. Simply increasing the number of parents may not safeguard the genetic diversity of a population when offspring of the additional parents have a low chance of being selected. The inheritance of selective advantage is crucial in the prediction of long-term genetic contributions, and thus for the prediction of inbreeding (Wray and Thompson 1990). Recently, Nomura (1997) studied inbreeding in open nucleus breeding systems with discrete generations, assuming that genetic contributions of parental groups (nucleus and commercial animals) to progeny remain unchanged after selection. As recognized by Nomura (1997), this is a critical assumption, and especially in populations with overlapping generations it is likely to be strongly violated.

Asymptotically, response from conventional gene flow is equal to response obtained using the well-known result of Rendel and Robertson (1950; Hill 1974). When gain obtained from conventional and modified gene flow was compared to simulation results, predic-

tions from both methods showed similar accuracy. For the prediction of genetic gain, the ratio of selection differential over generation interval is crucial, rather than the definitions of selection differential and generation intervals separately. When generation interval is defined as the average age of parents of all offspring, and selection differential is defined as the deviation of selected parents from the overall mean, valid predictions for genetic gain are obtained (James 1977). Conventional gene flow, therefore, is a valid method for predicting genetic gain. The relevance of the current theory lies in predicting the development of pedigree, *i.e.*, of the origin and turnover rate of genes, and in predicting rates of inbreeding; it does not primarily predict genetic gain.

Johan A. M. Van Arendonk is gratefully acknowledged for encouraging and giving P.B. the opportunity to visit J.A.W., and for giving useful comments on this manuscript. One author (J.A.W.) gratefully acknowledges the Ministry of Agriculture, Fisheries and Food (United Kingdom) for financial support. Jack C. M. Dekkers is acknowledged for giving very useful comments on this manuscript. This research was financially supported by the Netherlands Technology Foundation and was coordinated by the Life Sciences Foundation.

## LITERATURE CITED

Bichard, M., A. H. R. Pease, P. H. Swales and K. Özkütük, 1973 Selection in a population with overlapping generations. Anim. Prod. **17:** 215–227.

Bulmer, M. G., 1971 The effect of selection on genetic variability. Am. Nat. **105:** 201–211.

Cochran, W. G., 1951 Improvement by means of selection, pp. 449–470 in *Proceedings of the 2nd Berkeley Symposium on Mathematics, Statistics and Probability*, edited by J. Neyman. University of California Press, Berkeley.

Dempster, E. R., and I. M. Lerner, 1950 Heritability of threshold characters. Genetics **35:** 212–236.

Fisher, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. R. Soc. (Edinburgh), Trans. **52:** 321–341.

Hazel, L. N., 1943 The genetic basis for constructing selection indexes. Genetics **28:** 476–490.

Hill, W. G., 1974 Prediction and evaluation of response to selection with overlapping generations. Anim. Prod. **18:** 117–139.

James, J. W., 1977 A note on selection differential and generation length when generations overlap. Anim. Prod. **24:** 109–112.

James, J. W., 1987 Determination of optimal selection policies. J. Anim. Breed. Genet. **104:** 23–27.

James, J. W., and G. McBride, 1958 The spread of genes by natural and artificial selection in a closed poultry flock. J. Genet. **56:** 55–62.

Nomura, T., 1997 Prediction of effective population size in open nucleus breeding systems. J. Anim. Breed. Genet. **114:** 333–347.

Pearson, K., 1903 On the influence of natural selection on the variability and correlation of organs. Philos. Trans. R. Soc. Lond. A Math. Phys. Sci. **200:** 1–66.

Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, 1992 *Numerical Recipes in FORTRAN: The Art of Scientific Computing.* Cambridge University Press, Cambridge, UK.

Rendel, J. M., and A. Robertson, 1950 Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. J. Genet. **50:** 1–8.

Villanueva, B., and J. A. W. Woolliams, 1997 Optimisation of breeding programmes under index selection and constrained inbreeding. Genet. Res. **69:** 145–158.

Villanueva, B., N. R. Wray and R. Thompson, 1993 Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. Anim. Prod. **57:** 1–13.

Woolliams, J. A., 1998 A recipe for the design of breeding schemes. Proceedings of the 6th World Congress on Genetics Applied to Livestock Production, Vol. 25, Univ. New England, Armidale, Australia, p. 427.

Woolliams, J. A., and R. Thompson, 1994 A theory of genetic contributions. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Vol. 19, Univ. Guelph, Guelph, Ontario, Canada, pp. 127–134.

Woolliams, J. A., N. R. Wray and R. Thompson, 1993 Prediction of long term contributions and inbreeding in populations undergoing mass selection. Genet. Res. **62:** 231–242.

Woolliams, J. A., P. Bijma and B. Villanueva, 1999 The expected development of pedigree and its impact on genetic gain. Genetics (in press).

Wray, N. R., and R. Thompson, 1990 Prediction of rates of inbreeding in selected populations. Genet. Res. **55:** 41–54.

Wray, N. R., J. A. Woolliams and R. Thompson, 1990 Methods for predicting rates of inbreeding in selected populations. Theor. Appl. Genet. **80:** 503–512.

Wray, N. R., J. A. Woolliams and R. Thompson, 1994 Prediction of rates of inbreeding in populations undergoing index selection. Theor. Appl. Genet. **87:** 878–892.

Communicating editor: T. Mackay

## APPENDIX A

**Derivation of Λ:** Elements $\lambda_{kl}$ are obtained as $\lambda_{kl} = p_k^{-1}\,{}^1\!/_2\tau_x b_{S_{kl}I_k}$, where ${}^1\!/_2\tau_x$ is the regression of the index of the offspring on the breeding value of the parent of sex $x$ (Wray *et al.* 1994), and $b_{S_{kl}I_k}$ is the regression of the selection score (selected or not selected, *i.e.*, $S = 1$ or 0) of the offspring on its index. In Equation 2, $\lambda_{kl}$ is expressed per selected offspring, whereas $S$ is an expression per selection candidate, the difference being on average a factor $p_k^{-1}$. Tau is obtained as $\tau_x = 2\,\mathrm{Cov}(A_x, I_{\mathrm{offsp}})/\mathrm{Var}(A_x)$ resulting in $\tau_m = b_3$, $\tau_f = b_2\,(1 - 1/d) + b_3/d$. From a result of Robertson (appendix in Dempster and Lerner 1950), $b_{S_{kl}I_k} = p_k i_k \sigma_I^{-1}$. Resulting expressions are $\lambda_{kl} = {}^1\!/_2 b_3 i_k \sigma_I^{-1}$ for male parents and $\lambda_{kl} = {}^1\!/_2(b_2(1 - 1/d) + b_3/d) i_k \sigma_I^{-1}$ for female parents.

**Derivation of Π:** Elements $\pi_{kl}$ are obtained as $\pi_{kl} = \mathrm{Cov}(A_{i(l)}, A_{j(k)}{}^*/\mathrm{Var}(A_{i(l)}){}^*$, where * denotes (co)variances after selection of the offspring. Using Cochran (1951), $\sigma^*_{AB} = \sigma_{AB} - (\sigma_{AI}\sigma_{BI}/\sigma_I^2)\kappa$ for the calculation of $\mathrm{Cov}(A,B)$ after selection on $I$ gives

$$\mathrm{Cov}(A_{i(l)}, A_{j(k)})^* = \mathrm{Cov}(A_{i(l)}, A_{j(k)}) - \frac{\mathrm{Cov}(A_{i(l)}, I_{j(k)})\mathrm{Cov}(A_{j(k)}, I_{j(k)})}{\mathrm{Var}(I_{j(k)})}\,\kappa_k,$$

where $\mathrm{Cov}(A_{i(l)}, A_{j(k)}) = {}^1\!/_2\sigma^{2*}_{A(x)l}$, $\mathrm{Cov}(A_{i(l)}, I_{j(k)}) = {}^1\!/_2\tau_x\sigma^{2*}_{A(x)l}$, $\mathrm{Cov}(A_{j(k)}, I_{j(k)}) = \rho\sigma_A\sigma_I$, $\mathrm{Var}(I_{j(k)}) = \sigma_I^2$, $\sigma^{2*}_{A(x)l}$ is the additive genetic variance among selected parents, and $x$ denotes the sex of parent $i$. Assuming that $\mathrm{Var}(A_{i(l)})$ is little affected by selection of the offspring, *i.e.*, $\mathrm{Var}(A_{i(l)})^* = \sigma^{2*}_{A(x)l}$, the resulting expression becomes $\pi_{kl} = {}^1\!/_2(1 - \kappa_k\tau_x\rho\sigma_A\sigma_I^{-1})$.

## APPENDIX B

**Derivation of ΔG:** Genetic gain is obtained from $E(\Delta G) = \sum_{l=1}^{2c_{\max}} n_l E[r_{i(l)}a_{i(l)}]$ (Woolliams and Thompson 1994), where $E[r_{i(l)}a_{i(l)}]$ is the expectation of $r_{i(l)}a_{i(l)}$ among

selected individuals in category $l$. With $E[r_{i(l)}\, a_{i(l)}] = E[(u_{i(l)} + e_{i(l)})a_{i(l)}] = E[u_{i(l)}a_{i(l)}]$, it follows that $E[r_{i(l)}\, a_{i(l)}] = \alpha_l E[a_{i(l)}] + \beta_l E[a_{i(l)}(A_{i(l)} - \bar{A}_l)]$, where expectations are conditional on selection in category $l$. Furthermore,

$$E[a_{i(l)}] = \frac{\mathrm{Cov}(a_{i(l)}, I_{i(l)})}{\sigma_I^2} i_l \sigma_I = \frac{1}{2}\tau_w \sigma_{A0}^2 i_l \sigma_I^{-1},$$

where $\tau_w$ is the regression of the index on the Mendelian sampling effect, $\tau_w = b_1(1 - 1/n_o) + b_2(1/n_o - 1/n_o d) + b_3/n_o d$. With

$$\begin{aligned} E(A_{i(l)} - \bar{A}_l] &= 0,\ E[a_{i(l)}(A_{i(l)} - \bar{A}_l)] \\ &= \mathrm{Cov}[a_{i(l)},\ (A_{i(l)} - \bar{A}_l)]^* \\ &= \mathrm{Cov}[a_{i(l)},\ (A_{i(l)} - \bar{A}_l)] \\ &\quad - \frac{\mathrm{Cov}(a_{i(l)}, I_{i(l)})\mathrm{Cov}[(A_{i(l)} - \bar{A}_l), I_{i(l)}]}{\sigma_I^2}\kappa_l \\ &= \frac{1}{2}\sigma_{A0}^2(1 - \kappa_l \tau_w \rho \sigma_A \sigma_I^{-1}). \end{aligned}$$

Summing elements over categories, the resulting expression for genetic gain becomes

$$E(\Delta G_t) = \tfrac{1}{2}\sigma_{A0}^2\ [\tau_w \sigma_I^{-1} \sum_{l=1}^{2c_{\max}} n_l \alpha_l i_l + \sum_{l=1}^{2c_{\max}} n_l \beta_l (1 - \kappa_l \tau_w \rho \sigma_A \sigma_I^{-1})]_t.$$

## APPENDIX C

**Example for mass selection:** Consider a mass selection scheme ($b_1 = b_2 = b_3 = 1$) with three age classes, $\mathbf{N} = \mathrm{diag}\{20,0,0,10,0,30\}$, $h_0^2 = 0.4$, and $n_o = 8$. Selected proportions, selection intensities, and variance reduction coefficients are $p_1 = 0.1250$, $p_4 = 0.0625$, $p_6 = 0.1875$, $i_1 = 1.6467$, $i_4 = 1.9668$, $i_6 = 1.4357$, $\kappa_1 = 0.8171$, $\kappa_4 = 0.8504$, $\kappa_6 = 0.7877$. Tau equals (see equations in appendixes a and b): $\tau_w = \tau_m = \tau_f = 1$. The conventional gene flow matrix equals

$$\mathbf{G}_0 = \begin{bmatrix} 0.5 & 0 & 0 & 0.125 & 0 & 0.375 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.125 & 0 & 0.375 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Contributions and generation intervals from conventional gene flow are $n_1\alpha_1 = 0.2857$, $n_4\alpha_4 = 0.0714$, $n_6\alpha_6 = 0.2143$, $L = 1/\Sigma\alpha = 1.75$. Equations in the *deterministic prediction procedure* section were iterated until equilibrium variances were reached, resulting in $\sigma_{A(m)}^2 = 0.0630$, $\sigma_{A(f)}^2 = 0.1013$, $\sigma_A^2 = 0.3643$, $\sigma_I^2 = 0.9643$, $\rho = 0.3778$. Based on equilibrium variances, $\mathbf{G}$, $\Lambda$, $\Pi$, and $\mathbf{D}$ are

$$\mathbf{G} = \begin{bmatrix} 0.5 & 0 & 0 & 0.205 & 0 & 0.295 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.223 & 0 & 0.277 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0.193 & 0 & 0.307 \end{bmatrix},$$

$$\Lambda = \begin{bmatrix} 0.838 & 0 & 0 & 0.838 & 0 & 0.838 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1.001 & 0 & 0 & 1.001 & 0 & 1.001 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.731 & 0 & 0 & 0.731 & 0 & 0.731 \end{bmatrix},$$

$$\Pi = \begin{bmatrix} 0.346 & 0 & 0 & 0.346 & 0 & 0.346 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.339 & 0 & 0 & 0.339 & 0 & 0.339 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0.351 & 0 & 0 & 0.351 & 0 & 0.351 \end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0.187 & 0 & -0.130 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.171 & 0 & -0.138 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.198 & 0 & -0.124 \end{bmatrix}.$$

Categories without parents are not relevant, and have zeroes. $\mathbf{G}$ identifies the origin of selected offspring; *e.g.*, $g_{14} = 0.205$ means that a proportion of $2 \times 0.205 = 0.410$ of the selected 1-yr-old males (category 1) descends from 1-yr-old dams (category 4), *i.e.*, were born when their dam was 1 yr old. From $\mathbf{G}$, the generation interval based on selected offspring equals $L_1 = \frac{1}{2}\{0.5 + 0.205 + 3 \times 0.295\} + \frac{1}{2}\{10/40 \times (0.5 + 0.223 + 3 \times 0.277) + 30/40 \times (0.5 + 0.193 + 3 \times 0.307)\} = 1.595$. $\mathbf{D}$ represents the breeding value of selected subgroups as deviation from the total selected group, *e.g.*, $d_{46} = -0.138$ means that 1-yr-old selected females descending from 3-yr-old dams have an average breeding value of 0.138 units below the average of all selected 1-yr-old females.

Solutions from Method M were $(\mathbf{N}\alpha)^T = (0.312, 0, 0, 0.124, 0, 0.188)$, $\beta^T = (0.0201, 0, 0, 0.0161, 0, 0.0081)$, $L = 1.602$, $\Delta G = 0.3222$. Solutions from Method P were $(\mathbf{N}\alpha)^T = (0.336, 0, 0, 0.173, 0, 0.164)$, $\beta^T = (0.0220, 0, 0, 0.0182, 0, 0.0086)$, $L = 1.487$, $\Delta G = 0.3513$.